DOCUMENT RESUME

ED 398 277 TM 025 479

AUTHOR Linacre, John M.
TITLE Descriptive or Explanatory Measurement.
PUB DATE Apr 96
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Demography; Item Response Theory; *Literacy; *Mathematical Models; *Measurement Techniques; Prediction; Reading Comprehension; *Statistical Analysis
IDENTIFIERS *Descriptive Method; *Explanatory Analysis; National Adult Literacy Survey (NCES); Rasch Model

ABSTRACT

Measurement is prescriptive in that the model required to produce the numbers from the phenomena is known in advance. Statistical theory is descriptive because the model required to explain the numbers is inferred from the numbers and any other information the researcher can bring to bear. There are well-known families of measurement models and statistical models; combining them poses a number of problems, as is illustrated through an analysis of the relationship between literacy and the last grade level completed for 17,650 respondents to the National Adult Literacy Survey. One approach is to use measurement and then description by summarizing the literacy of adults by their last grade levels. A Rasch model approach can be used. Another approach, that of explanatory measurement, is to attempt to discover how much can be explained by last grade level and how much by demography. The two analyses are based on the same data but address different questions. If the sample used for the summary analysis adequately represents the country's population, then the results summarize the population's literacy as well. The explanatory analysis enables prediction of expected reading comprehension levels. The choice of analysis depends on whether the researcher wants a picture of the situation or insight into its formation. (Contains four figures and five references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
JOHN M. LINACRE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Descriptive or Explanatory Measurement

John M. Linacre
MESA Psychometric Laboratory, University of Chicago
E-mail: MESA@uchicago.edu, Tel. (312) 702-1596 FAX (312) 702-0248

ED 398 277

Measurement is concerned with representing phenomena by means of numbers with clearly defined properties. Most familiar and useful are numbers with equal-interval properties. Statistics is concerned with summarizing such numbers into a few parameter values. Measurement is prescriptive: the model required to produce the numbers from the phenomena is known in advance. Statistical theory is descriptive: the model required to explain the numbers is inferred from the numbers and any other information the researcher can bring to bear on the problem. There are well-known families of measurement models (Rasch models) and statistical models (regression models). Can these be combined into a one-step procedure?

Georg Rasch formulated his model as a measurement device for individuals, not as a means of performing demographic surveys (Rasch 1980). Accordingly he modelled each individual with one "ability" parameter, and each item with one "difficulty" parameter. Gerhard Fischer's linear logistic test model (LLTM, 1973) reconceptualizes each item parameter as a combination of factor parameters which explain the item's difficulty. This decomposition of item parameters has permitted useful research into what makes items hard (Smith et al. 1992).

Since the Rasch model is transparent to the difference between items and persons, it would seem that persons also could be "decomposed". This is done indirectly in Fischer's (1989) linear logistic models for change, in which a general change of person ability is reinterpreted as a general change in item difficulty. Direct decomposition of person ability parameters, however, can be easily accomplished. Nevertheless problems in selection, interpretation and communication arise.

Methods of Analysis

1) Descriptive measurement and explanatory analysis:

a) A two-facet Rasch analysis is performed in which each adult is measured and each item calibrated. Each adult obtains a measure and a standard error (which can be misfit inflated).

b) These measures are summarized by demographic factor (unadjusted for other factors).
This is the typical report displayed in newspapers.
Here problems of interpretation and communication arise.

c) Each adult is "decomposed" demographically according to gender, ethnicity, native language etc. This introduces the problem of decomposition selection. Many demographic variables are highly correlated.

d) The sizes of the demographic effects (and their standard errors) are estimated using regression. The procedure can be unweighted, S.E.-weighted (i.e., by effect sizes), or information-weighted.
This is the typical report displayed in research findings.

In summary,

response data = individual measures + stochasticity + misfit
individual measures = demographic effects + unexplained variance
unexplained variance = individual variance + measurement error

TM025479

BEST COPY AVAILABLE

2) Explanatory measurement:

a) Each adult is "decomposed" demographically according to gender, ethnicity, native language etc., i.e., each adult is treated as a random representative of the individual's demography.

b) A many-facet Rasch analysis is performed in which each demographic effect is measured. This introduces the problem of identifying the frame of reference. Though trivial mathematically, it presents difficulties in interpretation and communication.

c) The sizes of the demographic effects (and their standard errors) are estimated.
These reports are not yet in public circulation.

In summary, response data = demographic effects + stochasticity + misfit

The logit sizes of the demographic effects estimated by the two methods will differ. Since the stochasticity in method 2 comprises the stochasticity in method 1 plus the within-demographic-effect person variance, then method 2 will appear to be less discriminating. The logit difference between two items in method 2 will be smaller than in method 1.

Heteroscedascity of individuals within demographic cells will increase misfit and so distort the measurement of the effects. Only an estimate of the demographic effect mean is obtained directly; no estimate of within-demographic-effect sample variance is obtained. On the other hand, Bayesian imputation of measures to extreme scores is no longer required, because no meaningful demographic group will have an extreme score.

This study raises philosophical problems about "what we want to measure?" analogous to those raised in comparing raw scores and Rasch measures in answer to the question "how we want to measure?". Most audiences have been familiarized to raw scores, and feel uncomfortable with measures (despite their obvious advantages). Similarly, most audiences have been familiarized with descriptive measurement reports (usually at level 1b, sometimes at 1d). Are explanatory measurement reports (2c) an ideal to be aimed at, or a misleading distraction? The audience is encouraged to consider this question by comparing plots depicting different answers to the question "what size is the demographic effect?"

The NALS Literacy Survey

Consider an investigation of the relationship between literacy and last grade level completed. 24,944 subjects were interviewed for the 1992 National Adult Literacy Survey. Of these, 17,650 subjects have usefully complete test responses and demographic information. There were 184 literacy items on the Survey. Of these, 165 items provide interpretable literacy information.

Measurement then Description

One approach to the relationship between last grade level completed and adult literacy is to summarize the literacy of adults by last grade level. A Rasch analysis produces literacy measures for 17,650 adults on 165 items. The measurement model is

$$\log\left(\frac{P_{ni}}{1 - P_{ni}}\right) \equiv B_n - D_i$$

The 17,650 measures are then summarized by grade level. This description is a snap-shot of history. These summary descriptive measures are shown in the plot in Figure 1. These measures could then be subjected to the usual regression analysis. Since each measure has a standard error, a weighted regression could be performed. Weighting by inverse standard error (1/S.E.), rather than by information

(1/S.E.^{2}), is likely to be more useful (Wright 1994).

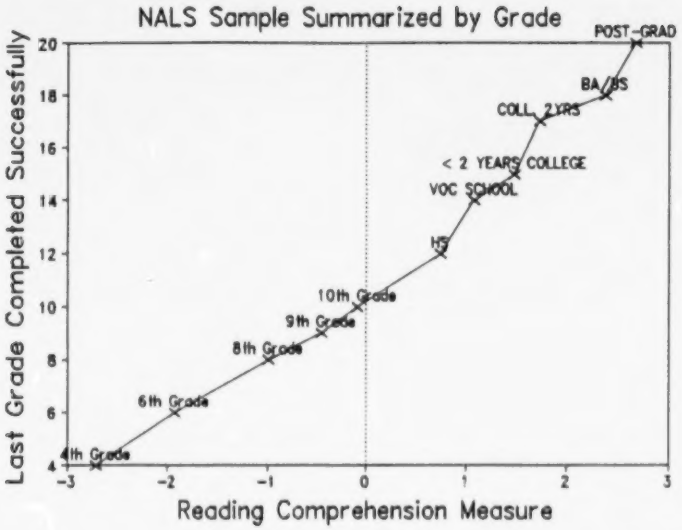

Figure 1. Reading Measures Summarized by Grade Level

Explanatory Measurement

Adult literacy is affected by many demographic and other factors. In an attempt to discover how much is explained by last grade level and how much by demography, an alternative demographic model is constructed. Now each respondent is thought of as a random member of each relevant demographic group. Each demographic group is parameterized to manifest one fixed effect. The demographies are last grade completed (G_g), personal income (M_m), first language (L_l), ethnicity (E_e), age (A_a) and sex (S_s). The explanatory measurement model is:

$$\log\left(\frac{P_{gmleasi}}{1 - P_{gmleasi}}\right) \equiv G_g + M_m + L_l + E_e + A_a + S_s - D_i$$

The grade effect measures this model produces are shown in the plot in Figure 2.

Why the differences?

Though the trend in the two plots are similar, there are clear differences. Figure 3 plots the two sets of grade level literacy measures. The relationship is close to linear, but not an identity line. The logit range for the summary measures is twice that for the explanatory measures. There are two reasons for this.

a) The length of the logit:

The measures are reported in logits, a probabilistically defined unit of measurement. The relationship between logits and the substantive definition of the latent variable depends on the amount of stochasticity in the data. The more noisy the data are, the more substantive advance along the latent variable there is per logit. The less noisy the data are, the larger (in logits) any given substantive difference appears. In perfect Guttman data, every respondent is infinitely far from every other respondent with a different raw score on the same items.

In the NALS example, the measurement model with 17,650 respondent parameters explains much more of the variance in the data, than the explanatory model with 54 demographic parameters. The unexplained variance, i.e., that within demographic groups, becomes part of the unexplained noise, inflating the measurement error. This larger measurement error reduces the logit range of the reported measures.

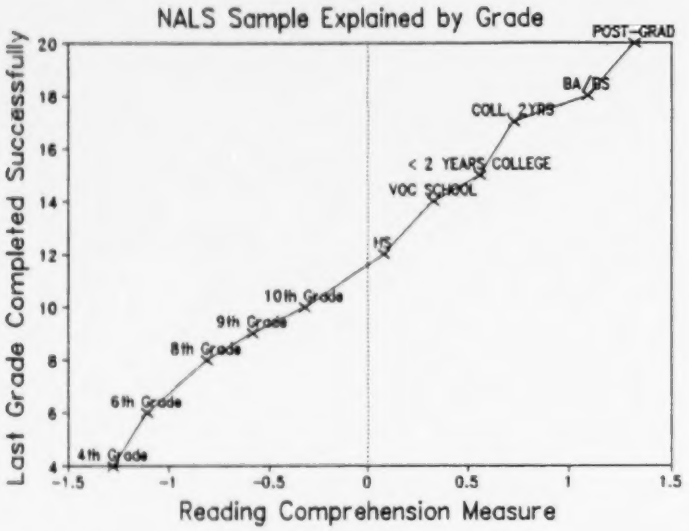

Figure 2. Explanatory Grade Measures.

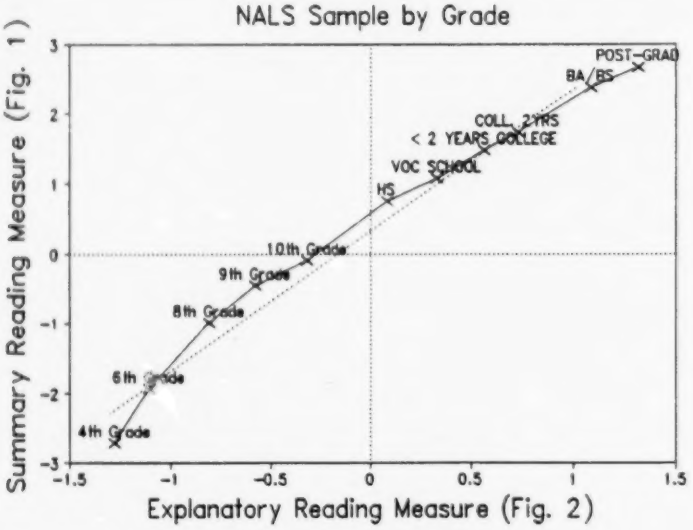

Figure 3. Comparison of Grade Level Measures.

The size of the logit change can be discovered by comparing the standard deviations of the item calibrations for the two analyses. In Figure 4, the standard deviation of the calibrations for the individual person measures is 1.2 times that for the explanatory model. This comparison of item calibrations equates the two alternative analyses by means of a part of the analysis that we can consider invariant.

b) Collinearity of explanatory variables:
Higher reading comprehension would be expected of those respondents with more education, but also of those respondents with higher income. A likely contributing factors to higher income are both more education and more reading comprehension. The explanatory model partitions the reading comprehension

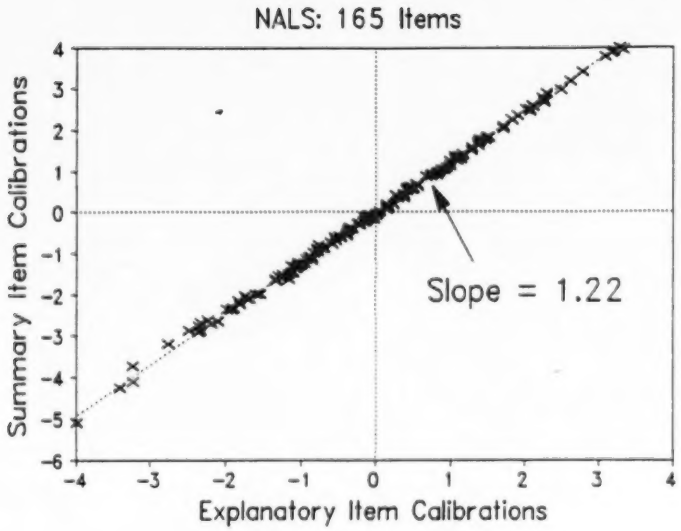

Figure 4. Comparison of item calibrations.

effect of the demographic variables for this sample. According to Figure 3, additional education by itself (according to the explanatory run) accounts for about half the increase in literacy achieved by those with more education. After adjusting for the change in logit size, 63% of the increase is due to additional education.

Descriptive or Explanatory Measurement?

These two analyses, though based on the same data, answer different questions. The summary analysis reports the reading comprehension associated with each last grade level according to this sample. If this sample adequately represents the US population, then these results also summarize the population. The explanatory analysis enables us to predict, for any people similar to those sampled, from their demographic characteristics what are their expected reading comprehension levels. The concave form of Figure 3 indicates that the summary statistics underemphasize the close relationship between reading comprehension and successful completion of higher education. The choice of analysis depends on whether the researcher desires a snapshot of the current situation or insight into its formation.

Fischer G (1973) The linear logistic test model as an instrument in educational research. Acta Psychologica 36, 207-220.

Fischer G (1989) An IRT-based model for dichotomous longitudinal data. Psychometrika 54, 599-624.

Smith RM, Kramer GA, Kubiak AT (1992) Components of Difficulty in Spatial Ability Items, In Wilson M. (Ed.) Objective Measurement: Theory into Practice. Vol. 1.

Rasch G (1980) Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press.

Wright BD (1994) Part-test vs. whole-test measures. Rasch Measurement Transactions 8:3 p. 376-366.